



TITLE:

Relationship between job opportunities and economic environments measured from data in internet job searching sites

AUTHOR(S):

Sato, Aki-Hiro; Shimizu, Chihiro; Mizuno, Takayuki; Ohnishi, Takaaki; Watanabe, Tsutomu

CITATION:

Sato, Aki-Hiro ...[et al]. Relationship between job opportunities and economic environments measured from data in internet job searching sites. *Procedia Computer Science* 2015, 60: 1255-1262

ISSUE DATE:

2015-09-01

URL:

<http://hdl.handle.net/2433/224812>

RIGHT:

Copyright © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 60 (2015) 1255 – 1262

Procedia
Computer Science

19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Relationship between job opportunities and economic environments
measured from data in internet job searching sitesAki-Hiro Sato^{*a}, Chihiro Shimizu^b, Takayuki Mizuno^c, Takaaki Ohnishi^d,
Tsutomu Watanabe^e^aGraduate School of Informatics, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501 Japan^bInstitute of Real Estate Studies, National University of Singapore, 21 Heng Mui Keng Terrace, Singapore 119613^cNational Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Japan^dGraduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033 Japan^eGraduate School of Economics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033 Japan

Abstract

This study investigates the number of job opportunities collected from a Japanese job searching site (“fromA navi”) and an international job searching site (“Indeed”). We confirm that a relationship between the number of job opportunities and socioeconomic quantities (the population, the numbers of firms and workers) in each 1-km grid square in Japan. The number of workers is the best explanatory variable to explain the number of job opportunities in Japan. The regression coefficients can be used as an indicator to grasp Japanese macroeconomic conditions. From a global point of view, we analyse the number of job opportunities in about 16,000 cities all over the world. We confirm the daily number of job opportunities in each city varies in time and show some associations with macroeconomic indicators. We compute a relationship between means of the daily number of job opportunities and their standard deviations and confirm that it follows a scaling relationship with power law exponent $\alpha = 1$. A possible model based on Poisson processes with intensity of which varies in time on the basis of a common noise is proposed to explain the phenomenon empirically observed.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of KES International

Keywords: Job opportunities; regression analysis; governmental grid square statistics data; fluctuation scaling

1. Introduction

Due to development of Information and Communications Technology, rich data on socioeconomic systems are available. In recent years, Web information is expected to be a useful resource to analyse regional dependence of

^{*} Aki-Hiro Sato Graduate School of Informatics, Kyoto University. Tel.: +81-75-753-5515 ; fax: +81-75-753-5515.
E-mail address: aki@i.kyoto-u.ac.jp

socioeconomic conditions and several data-driven applications such as flu epidemics detection¹ and financial market prediction² have been proposed.

Recently, Japan faces the challenges of an ageing society. About 25% of the total population in Japan exceeds the age of 65, which counts over 30 million.³ Specifically, in an ageing society, we need to consider a balance between demand and supply in both economic and social dimensions. This implies that it is important to understand regional dependency between population and workforce. We will need to consider the population balances from a social sustainability point of view. In this study, we investigate a relationship between the number of job opportunities⁴ and socioeconomic quantities (the population, the numbers of firms and workers) provided by e-Stat, which is a portal site handled by Statistics Bureau, Ministry of Internal Affairs and Communications. These socioeconomic quantities are served as governmental open data⁵ and recorded as 1-km grid square statistics. Moreover, we examine the number of job opportunities at about 16,000 cities in 50 countries, which were collected from an Internet job searching site.⁶ These form multi-dimensional time series generated from counting processes.

Janet Yellen, who is the Chair of the Board of Governors of the Federal Reserve System, watches an array of gauges to steer the economy toward price stability and maximum sustainable employment. The most important indicators which she is watching are known as Janet Yellen's Dashboard. In fact, there are several macroeconomic indicators related to employment, such as unemployment rates, employment to population ratio, valid job vacancy rate and so on.

The aim of this paper is to quantify dependence of job creations on macroeconomic conditions. In fact, both sampling frequency and spatial resolution of employment indicators reported based on governmental statistics are low. Thus, there may be a room to improve employment indicators combining data from Internet sites and data from governmental statistics survey.

In this study, we use governmental grid square statistics data to measure macroeconomic conditions and the number of job opportunities collected from an Internet job searching site. On the one hand, the advantage of governmental statistics is both high accuracy and large coverage, however, their disadvantage is a low sampling frequency. On the other hand, the advantage of Internet data is a high sampling period, however, their disadvantage is low coverage. We attempt to adjust data obtained from an Internet site with a high sampling period by using governmental statistics data with high accuracy whilst a low sampling frequency.

2. Empirical analysis of job opportunities

2.1. Japanese cases ("fromA navi" data)

We collect raw data on job opportunities via a Web API provided by "fromA navi" in Recruit Web Service.⁴ The data contain job types, duration of calls, wages, geographical positions (latitude and longitude) of job interview places and/or working places, prefecture, the name of firms calling the job opportunities and contents of job opportunities.

From this data, we calculated 1-km grid square statistics on the number of job opportunities. We utilised JISX0410 of Japanese Industrial Standards in order to calculate 1-km grid square code. The 1-km grid square code is computed as

$$\text{grid square code} = puqvrw, \quad (1)$$

where

$$\begin{cases} \lfloor \text{latitude} \times 60 \div 40 \rfloor = p \text{ (} p \text{ is two digits.)}, & a = (\text{latitude} \times 60 \div 40 - p) \times 40, \\ \lfloor a \div 5 \rfloor = q \text{ (} q \text{ is one digit.)}, & b = (a \div 5 - q) \times 5, \\ \lfloor b \times 60 \div 30 \rfloor = r \text{ (} r \text{ is one digit.)}, & c = (b \times 60 \div 30 - r) \times 30, \\ \lfloor \text{longitude} - 100 \rfloor = u \text{ (} u \text{ is two digits.)}, & f = \text{longitude} - 100 - u, \\ \lfloor f \times 60 \div 7.5 \rfloor = v \text{ (} v \text{ is one digit.)}, & g = (f \times 60 \div 7.5 - v) \times 7.5, \\ \lfloor g \times 60 \div 45 \rfloor = w \text{ (} w \text{ is one digit.)}, & h = (g \times 60 \div 45 - w) \times 45, \end{cases} \quad (2)$$

By using geographical information of working places, we computed the daily number of job opportunities for each grid square. Job opportunities all over Japan highly dense in large cities such as Tokyo and Osaka. Several mega cities (Tokyo, Osaka, Nagoya, Sapporo and Fukuoka) keep a large number of job opportunities.

It can be hypothesised that the number of job creations is related to the population and production in each place. To verify such a hypothesis, we conduct regression analysis of the number of job opportunities with respect to three socioeconomic quantities. We set three types of explanatory variables such as the population, the numbers of firms and workers. Three types of 1-km grid square statistics data selected as explanatory variables were downloaded from e-Stat, which is a portal site to provide governmental open data.⁵ Specifically, we used 1-km grid square statistics data of the census population in 2010 (reported by Statistics Bureau, Ministry of Internal Affairs and Communications), of the numbers of firms and workers in 2012 (reported by Ministry of Economics, Trade and Industry). Finally, we constructed a database by linking the governmental grid square statistics data and the number of job opportunities by using grid square codes.

First of all, we examine correlations among the candidates of explanatory variables. We extracted 126,507 grid squares where any values in three variables are not zero from the 1-km governmental grid square statistics data.

Table 1 shows correlation coefficients (Pearson, Kendall, and Spearman correlation coefficients) among the population, the numbers of firms and workers in each grid square. We confirmed that a positive correlation among three quantities. Specifically, the correlation between the numbers of firms and workers is the best in three pairs. The next largest correlation coefficient is obtained in a pair between the population and the number of workers. This implies that the population shows the weakest correlation with the number of firms in three pairs and that the correlation between the numbers of firms and workers is the strongest in three pairs.

Table 1. Correlations among the population, the total number of firms and the number of workers. The number of grid squares is 126,507.

method	population - # of workers	population - # of firms	# of firms - # of workers
Pearson	0.645989	0.433767	0.883059
Kendall	0.715239	0.617294	0.790496
Spearman	0.876813	0.801737	0.926238

In this study, we assume the population, the number of firms and the number of workers as macroscopic variables to characterise job creation. We attempt to investigate an explanatory power of these quantities to the number of job opportunities. In general, if we use correlated variables as explanatory variables for multiple regression analysis, then we face the multicollinearity problem. In order to avoid the complexity due to the multicollinearity, we assume a simple relationship between the number of job opportunities and an explanatory variable selected from the three quantities and examine temporal development of regression coefficients.

Suppose $E(t)$ is a set of grid square codes where both the number of job opportunities and an explanatory variable selected from three quantities are not zero on observation date t . We express the number of grid squares included in $E(t)$ as $N(t)$. We attempt to find a relationship between the daily number of job opportunities and the governmental grid square statistics in each grid square. Figure 1 shows the double-logarithmic plots between the daily number of job opportunities and (a) the population, (b) the number of firms and (c) the number of workers in grid squares included in $E(t)$ on $t = 10$ March 2015.

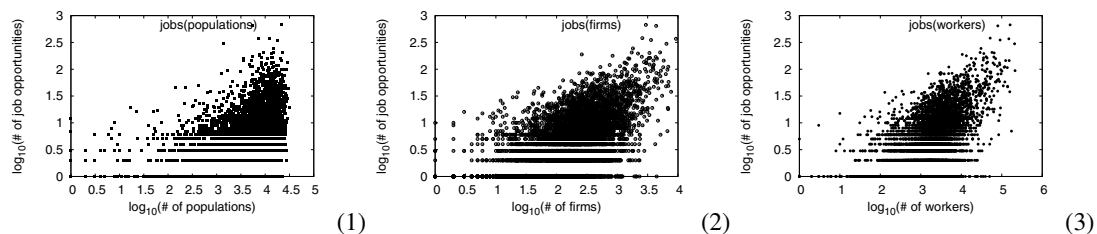


Fig. 1. Double-logarithmic scatter plots between the number of job opportunities and (1) the population ($N(t) = 7,298$), (2) the number of firms ($N(t) = 7,313$) and (3) the number of workers ($N(t) = 7,313$) on $t = 10$ March 2015.

The relationship between the number of job opportunities and the population is not clear, however, the relationships between the number of job opportunities and the number of firms, and the number of workers show some correlations. The number of job opportunities may be explained by using the numbers of firms and workers in each grid square more than by using the population.

According to Zhang and Yu (2010)⁷ and Bettencourt *et al.* (2007)⁸, there are allometric relationships between extensive macroscopic variables A and M in areas, cities or countries. This is described as a power-law relationship:

$$A = CM^\alpha, \quad (3)$$

where α is a power law exponent. If $\alpha > 1$, the property of X has a super-linear relation with respect to M . $\alpha < 1$ means that X has a sub-linear relation with respect to M . $\alpha \approx 1$ shows a linear relation between X and M . Therefore, we may assume power law relationships for explained variable $Y(c)$ and explanatory variable $X_i(c)$:

$$Y(c) = C_i X_i(c)^{\alpha_i}, \quad (4)$$

where $Y(c)$ represents the number of job opportunities in grid square $c \in E(t)$ and $X_i(c)$ is the population ($i = p$), the number of firms ($i = f$) and the number of workers ($i = w$) in grid square $c \in E(t)$. C_i and α_i are positive constants.

Under these assumptions, we compute regression coefficients for a logarithmic form for Eq. (4):

$$\log Y(c) = \log C_i + \alpha_i \log X_i(c). \quad (5)$$

Furthermore, to quantify goodness-of-fit for Eq. (5) we utilise adjusted R squared defined as

$$\text{adjusted } R_i^2(t) = 1 - \frac{\sum_{c \in E(t)} (\log Y(c) - \log C_i - \alpha_i \log X_i(c))^2 / (N - 2)}{\sum_{c \in E(t)}^N (\log Y(c) - \overline{\log Y})^2 / (N - 1)}, \quad (6)$$

where N is the number of grid squares included in $E(t)$ and

$$\overline{\log Y} = \frac{1}{N(t)} \sum_{c \in E(t)} \log Y(c). \quad (7)$$

A value of the adjusted R squared expresses an explanatory power of an explanatory variable for an explained variable. If adjusted R squared is 1, then it implies a perfect linear relation. If adjusted R squared is 0, then there is no relation between the explained and explanatory variables. A large value of the adjusted R squared is related with a large explanatory power of an explanatory variable against an explained variable.

Table 2 shows parameter estimates (α_i and $\log C_i$), their t -values and p -values and a value of adjusted R squared for Eq. (4) between the number of job opportunities on 10 March 2015 and three explanatory variables (the population, the numbers of firms and workers). A value of adjusted R square for the population is smaller than other explanatory variables (the number of firms and workers). The values of adjusted R square for the numbers of firms and workers are almost equal. Since the t -value for the power law exponent α_i is sufficiently large and its p -value is sufficiently small ($< 2 \times 10^{-16}$). Therefore, the null hypothesis that the value of α_i is zero is rejected with 1% statistical significance level. This means that the numbers of firms and workers have an explanatory power to the number of job opportunities better than the population with a statistical significance.

Table 2. Parameter estimates (α_i and $\log C_i$) of a relationship between the number of job opportunities and one of three explanatory variables (the population, the numbers of firms and workers) on 10 March 2015.

explanatory variables	population ($i = p$)	firms ($i = f$)	workers ($i = w$)
The number of grid squares	7,298	7,313	7,313
α_i	0.376	0.473	0.446
t -value	42.337	55.033	56.004
p -value	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
$\log C_i$	-1.737	-1.063	-1.950
t -value	-23.79	-23.84	-32.88
p -value	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$
Adjusted R squared	0.197	0.298	0.300

Figure 2 (left) shows adjusted R squared for α_p , α_f and α_w in terms of sampling date during the period from 25 February to 19 May 2015. It is confirmed that the adjusted R squared of Eq. (4) for the population are smaller than values for the number of firms and the number of workers. The adjusted R squared for the numbers of firms

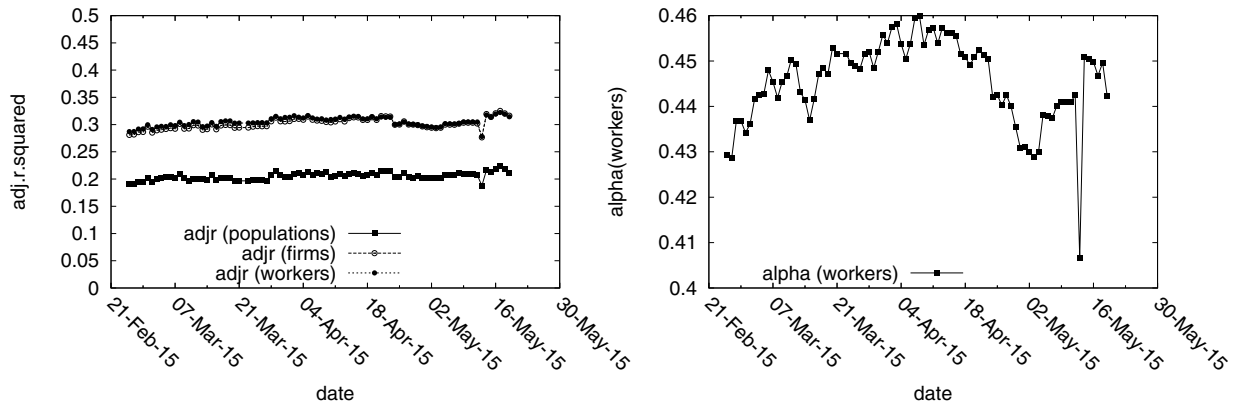


Fig. 2. (left) Trends in adjusted R squared of a regression between the number of job opportunities and the population (filled squares), the numbers of firms (unfilled circles) and workers (filled circles) and (right) temporal development of power-law exponent α_w during the period from 22 February to 19 May 2015.

and workers are almost same as each other. This is consistent with correlation coefficients among three explanatory variables shown in Tab. 1. This implies that an explanatory power of Eq. (4) for the population is less than the numbers of firms and workers.

Regression coefficients α_p , α_f and α_w may be used as indicators to measure a job creation rate in Japan. The t -values of regression coefficients α_f and α_w are also sufficiently large during the observation period and the null hypothesis that α_i is zero is rejected for three cases with 1% statistical significance. Thus, the density of job creations is correlated with the density of firms and workers which have been already realised. The values of adjusted R squared are stable for three cases during the observation period. This means that explanatory powers of three quantities (the population, the numbers of firms and workers) show weak seasonal dependence.

Moreover, Fig. 2 (right) shows temporal development of power-law exponent α_w against the number of workers. The value of α_w increased from the end of February 2015 eventually and decreased the end of April 2015. The reason why the value of α_w is low during the period may be because this period corresponded to Golden week holidays in Japan. Many firms took holidays in this period and business activities during the period were not so high. Generally speaking, job opportunities can be assumed to increase before business activities increase. The temporal development of power-law exponent α_w shown in Fig. 2 (right) may be explained by this assumption.

2.2. Global job markets (“Indeed” data)

Next, we move to a global trends of labour market. We collected the number of job opportunities in about 16,000 cities from a Web API provided by “Indeed”.⁶ From this data, we can observe the daily number of job opportunities in each city from a global point of view. Figure 3 shows the daily number of job opportunities in selected 15 cities. In New York, USA, we confirmed that the number of opportunities decreased from August to September 2013 and that from January 2014 the number increased. This is associated with socioeconomic conditions of United States. The number of job opportunities forms multi-dimensional time series described as counting processes. According to a theory of fluctuation scaling,^{9,10,11} we have a scaling relationship between sample means and sample standard deviations.

Suppose that $x_i(t)$ denotes the number of job opportunities in city i . The scaling-relationship

$$\sigma_i = C m_i^\alpha \quad (8)$$

where m_i and σ_i are, respectively, sample mean and sample standard deviation, which are defined as

$$m_i = \frac{1}{T} \sum_{t=1}^T x_i(t), \quad \sigma_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_i(t) - m_i)^2}. \quad (9)$$

C is a positive constant and α takes a value ranging from $1/2$ to 1 .

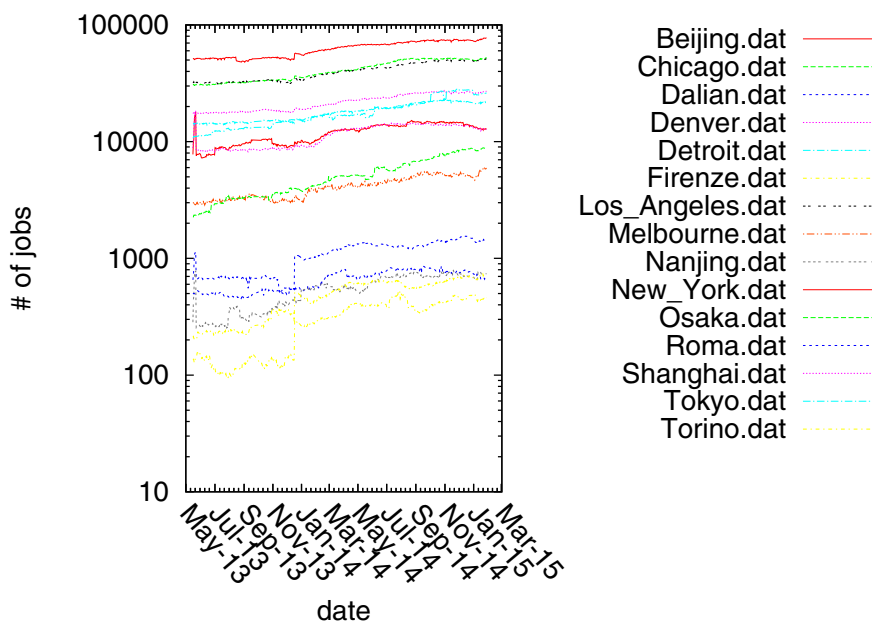


Fig. 3. The number of job opportunities in selected 15 cities (Beijing, China; Chicago, USA; Dalian, China; Denver, USA; Detroit, USA; Firenze, Italy; Los Angeles, USA; Melbourne, Australia; Nanjing, China; New York, USA; Osaka, Japan; Roma, Italy; Shanghai, China; Tokyo, Japan; Torino, Italy) during the period from 16 May 2013 to 29 January 2015.

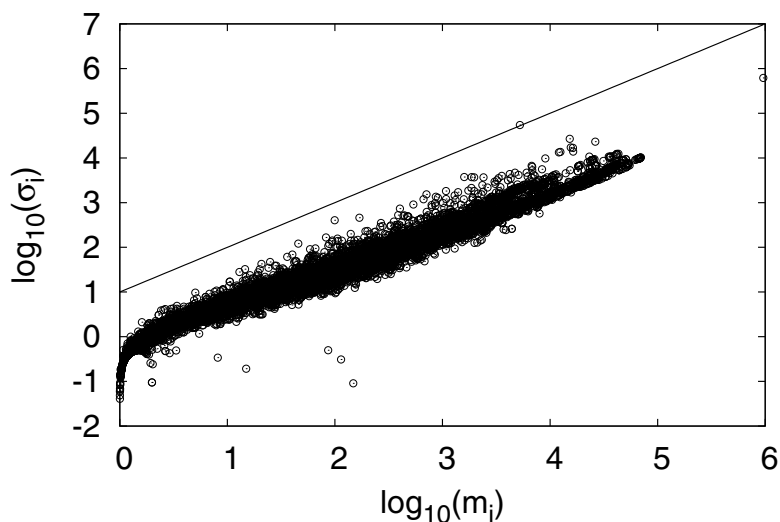


Fig. 4. A relationship between sample means computed from the numbers of job opportunities and their sample standard deviations. The duration is from 16 May 2013 to 29 January 2015.

The power law exponent α takes a value ranging from $1/2$ to 1 . If $\alpha = 1/2$ then processes $x_i(t)$ are mutually independent Poisson processes. If $\alpha = 1$ then $x_i(t)$ are correlated and their intensities are driven by a same factors.

Figure 4 shows double-logarithmic scatter plots of m_i and σ_i computed from the actual time series during the period from 16 May 2013 to 29 January 2015. We ignored time series of cities for $T < 100$ when we compute sample means and sample standard deviations.

We confirm that there is a clear power law relationship between sample means and sample standard deviations. The slope is fitted as $\alpha = 0.5$ for $m_i < 148.4$ and $\alpha = 1.0$ for $m_i > 148.4$. This means that the daily number of job opportunities can be modelled as Poisson processes with an intensity of which varies in time by a common stochastic variables. This phenomena can be described as Poisson processes of which intensity fluctuates with a common noise source (See Appendix Appendix A). However, the scatter plots between means and standard deviations shown in Fig. 4 do not form a simple curve which is described by the stochastic model introduced in Appendix A. This may result from some groups with strong correlations. We need further investigations to understand regional dependence of global job markets and associations with macroeconomic indicators.

3. Conclusions

We collected data on Japanese job opportunities from a Japanese job searching site. We examined a relationship between the number of job opportunities and socioeconomic quantities (the population, the numbers of firms and workers) using 1-km grid square statistics data. We conducted regression analysis of the number of job opportunities in each 1-km grid square with respect to one of three socioeconomic quantities under the assumption of their power-law relationship. We confirmed that the number of workers is the best explanatory variable in the three quantities and that the power-law exponent varied in time depending on the situations of job conditions. The power-law exponent between the number of job opportunities and the number of workers in each grid square may be used as an indicator to quantify job creation ratio.

Furthermore, we collected data on job opportunities in 50 countries from an international job search site. We counted the number of job opportunities in about 16,000 cities. The numbers varied in time and were expected to be associated with macroeconomic conditions of cities.

As future work, we need to conduct regression analysis of the number of job opportunities in terms of other socioeconomic quantities. Moreover, we should confirm a correlation between power-law exponent α_i and macroeconomic indicators such as interest rates and gross domestic product (GDP). We further need to compute correlations between the number of job opportunities and average wage in each 1-km grid square and to identify group structure where the numbers of job opportunities are correlated with one another. We also need to investigate regional dependence of wages and dependence of wages on job classes. This analysis will contribute to understand a balance between demand and supply in labour markets from a global point of view.

Appendix A. A fundamental theory of fluctuation scaling

A.1. Case of $\alpha = 1/2$

Suppose that $x_i(t)$ are integer random variables following mutually independent Poisson processes. Namely, $x_i(t)$ is sampled from a Poisson distribution with intensity $k_i\lambda$,

$$\Pr(x_i; k_i\lambda) = e^{-k_i\lambda} \frac{(k_i\lambda)^{x_i}}{x_i!}, \quad (\text{A.1})$$

where k_i is a positive constant depending of site i . Since mean of a Poisson distribution in Eq. (A.1) is $m_i = k_i\lambda$ and the standard deviation is $\sigma_i = k_i^{1/2}\lambda^{1/2}$, we have a power-law relationship between the means and standard deviations with $\alpha = 1/2$ such that $\sigma_i = m_i^{1/2}$.

A.2. Case of $\alpha = 1$

Suppose that $x_i(t)$ are integer random variables following Poisson processes with intensity k_i , where k_i is driven by a same random force. The intensity of the simplest model is assumed to be described as $k_i\lambda$, where k_i is a positive

constant depending on site i and λ is a uniform random variable ranging from 0 to d . The mean and standard deviation are, respectively, calculated from

$$\begin{aligned} m_i &= \int_0^d \sum_{x_i=1}^{\infty} x_i e^{-k_i \lambda} \frac{(k_i \lambda)^{x_i}}{x_i!} d\lambda \\ &= \int_0^d k_i \lambda d\lambda = \frac{k_i d^2}{2}, \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} \sigma_i^2 &= \int_0^d \sum_{x_i=1}^{\infty} x_i^2 e^{-k_i \lambda} \frac{(k_i \lambda)^{x_i}}{x_i!} d\lambda - m_i^2 \\ &= \int_0^d (k_i \lambda + k_i^2 \lambda^2) d\lambda - m_i^2 \\ &= \frac{k_i}{2} d^2 + \frac{k_i^2}{3} d^3 - m_i^2 = m_i \left\{ m_i \left(\frac{4}{3d} - 1 \right) + 1 \right\}. \end{aligned} \quad (\text{A.3})$$

Equation (A.3) can be approximated as $\sigma_i \propto m_i^{1/2}$ for $m_i \ll 1$ and $\sigma_i \propto m_i$ for $m_i \gg 1$. Namely, we have a power-law relationship between m_i and σ_i with $\alpha = 1$.

References

1. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS. Detecting influenza epidemics using search engine query data. *Nature*, 2009; **457**(19):10121014.
2. Blasco N, Corredor P, Del Rio C, Santamaria R. Bad news and Dow Jones make the Spanish stocks go round. *Eur J Oper Res*, 2005; **163**(1):253275.
3. Ministry of Internal Affairs and Communications in Japan.
4. “fromA navi” in Recruit Web Service, <http://www.froma.com/>
5. e-Stat, Statistical Bureau, Ministry of Internal Affairs and Communications in Japan, <http://www.e-stat.go.jp/>
6. “Indeed”, <http://jp.indeed.com/>
7. Zhang J, Yu T. Allometric scaling of countries. *Physica A*, 2010; **389**:4887–4896.
8. Bettencourt LMA, Lobo J, Helbing D, Kühnert C, West GB. Growth, innovation, scaling, and the pace of life in cities. *Proc. Nat. Acad. Sci. USA*, 2007; **104**:7301–7306.
9. Argollo de Menezes M, Barabasi A-L. Separating the internal and external dynamics of complex systems. *Phys Rev Lett*, 2004; **92**:028701
10. Eisler Z, Bartos I, Kertész J. Fluctuation scaling in complex systems: Taylor’s law and beyond. *Advances in Physics*, 2008; **57**:89-142
11. Sato A-H, Hayashi T, Holyst JA. Comprehensive Analysis of Market Conditions in the Foreign Exchange Market: Fluctuation Scaling and Variance-Covariance Matrix, *Journal of Economic Interaction and Coordination*, 2012; **7**:167–179